



L'homme multidimensionnel et l'opinion publique, quelques réflexions sur les statistiques et les sciences sociales

Nicolas Bouleau

► To cite this version:

Nicolas Bouleau. L'homme multidimensionnel et l'opinion publique, quelques réflexions sur les statistiques et les sciences sociales. Journées académiques avril 2004, Apr 2004, Lille, France. halshs-00781620

HAL Id: halshs-00781620

<https://shs.hal.science/halshs-00781620>

Submitted on 28 Jan 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'HOMME MULTIDIMENSIONNEL ET L'OPINION PUBLIQUE, REFLEXIONS SUR LES STATISTIQUES ET LES SCIENCES SOCIALES

Nicolas Bouleau, Lille, IREM, Avril 2004

Résumé

Nous abordons trois thèmes. D'abord la question du hasard dans le champ sémantique. Il est facile de formaliser le hasard à propos d'objets, les boules des urnes, etc., mais plus délicat de l'envisager dans le registre de ce qui fait sens. Quelques éléments historiques et l'analyse de cette difficulté nous conduisent à une sorte d'incompatibilité ou de principe d'exclusion.

Ensuite "l'opinion publique". Les sondages lorsqu'ils sont nombreux convergent-ils vers ce qu'on peut appeler l'opinion publique ? Un peu de mathématiques convainc qu'il n'y a là rien d'évident. Faut-il dès lors se tourner vers les enquêtes qualitatives très utilisées des sociologues ?

Enfin l'infinitude de l'individu. Les espaces vectoriels de dimension infinie ont, comme on sait, la particularité que la boule unité n'y est jamais compacte. La finitude des caractéristiques génétiques de l'homme permet-elle de le penser comme fini-dimensionnel ? Nous récusons cette vision réductionniste par une réflexion sur l'épigenèse.

Il est difficile de considérer comme fortuit ce qui nous concerne, notre existence même, ce qui agit sur nos motivations. Les lectures symboliques que nous faisons de notre environnement nous induisent à penser que notre comportement dans ce cadre de vie est, au moins partiellement, voulu. Pourtant les statisticiens et économètres apportent des chiffres, des proportions, des corrélations, des invariants et tendent à décrire le monde social comme un système qui suit des lois objectives, certes en évolution, mais dont on peut aussi appréhender les tendances. A la limite, de ce point de vue qui se veut scientifique, ce sont les subjectivités individuelles qui peuvent être considérées comme au hasard. La thèse de Jacques Monod de rapporter la biologie à la seule dualité du hasard et de la nécessité, épouse le même schéma.

Il y a pourtant un grand absent de cette vision néo-matérialiste influencée par l'exemple de la physique, c'est le langage. Le monde n'est pas fait uniquement de matière et de champs de forces, il y a l'écrit et la parole, les langages scientifiques avec leurs conventions, leurs transformations et leur étonnante puissance, qui viennent modifier le monde lui-même et cette vision par trop simpliste. Cela ouvre la voie à de nombreuses interrogations philosophiques. Nous allons tenter d'en présenter quelques-unes.

Le hasard et le champ sémantique

Plutôt que de rechercher une base doctrinale qui serait le point de départ de notre analyse, il est intéressant de suivre le trajet et les étonnements d'un grand esprit rationnel, reconnu comme tel, devant ce qu'il rencontre en toute bonne foi. Je veux parler de Ferdinand de Saussure (1857-1913) célèbre pour ses recherches philologiques sur l'origine des langues indo-européennes et en particulier son invention des "coefficients sonantiques" qui furent confirmés par les découvertes ultérieures des inscriptions hittites. Saussure fut aussi un grand professeur qui synthétisa son érudition encyclopédique en un cours remarquable de clarté, le

Cours de Linguistique Générale, qui sera vu comme un manifeste structuraliste par Barthes, Lacan, Lévi-Strauss et d'autres intellectuels après la seconde guerre mondiale.

Ferdinand de Saussure, et c'est moins connu, a passé une part considérable de son temps à des recherches sur la poésie latine qu'il a finalement tenues secrètes toute sa vie. Il a en couvert pas moins de 141 cahiers conservés à la bibliothèque de Genève. Son idée était que dans la poésie latine des 2^{ème} et 1^{er} siècles avant J.-C. (et aussi dans d'autres corpus latins, grecs ou germaniques) se trouvaient *des mots cachés* ou hypogrammes qui étaient des noms de Dieux ou de personnages historiques à qui le poème était dédié. Ces mots se trouvaient dans le poème sous forme de syllabes séparées que l'on entendait comme un accompagnement de la musique du poème. Il s'apprêtait à publier un traité sur le sujet. Mais avant de livrer cette accumulation de trouvailles au public, il hésita. Il demanda l'avis d'un de ses élèves sur la possibilité que les mots cachés ne soient simplement dus au hasard des combinaisons de sons. L'affaire se termina tristement. Pour en avoir le cœur net, Saussure s'attela à des poèmes en latin écrits par un de ses contemporains italiens et soumit à ce collègue des découvertes d'hypogrammes dans ses sonnets. Le collègue après un mot de politesse ne donna pas suite à la correspondance. Saussure crut s'être fourvoyé et ne publia rien¹.

Pour évaluer la pertinence de l'inquiétude de Saussure, il faut tenter de voir si on doit attribuer les trouvailles cachées à la volonté du poète ou au hasard. On a d'une part à calculer par un modèle mathématique de texte, la probabilité d'occurrence des hypogrammes en fonction du nombre de caractères, soit p_1 . D'autre part compter dans un corpus précis la fréquence des hypogrammes soit p_2 et conclure si $p_2 > p_1$ que le corpus est significativement doté d'hypogrammes.

Mais pour le premier calcul, il faut modéliser du texte qui respecte non seulement les fréquences des syllabes, mais celle des mots et des phrases, et même qui reflète la langue "régionale" dont est extrait le corpus étudié. Quant au second calcul, il est rendu imprécis par les tolérances que Saussure admet sur les sons et les emplacements. En y regardant de plus près on se convainc d'une difficulté tout à fait sérieuse due au fait que c'est *le sens* du poème qui fournit le champ des mots qui peuvent être considérés comme hypogrammes ou non.

Antoine Augustin Cournot (1808-1877), à la fois mathématicien, économiste et philosophe, avait déjà pointé une difficulté de cette nature dans ses réflexions sur les probabilités et en particulier dans sa critique de la notion d'*événement remarquable*, présentée sans précaution dans un traité du probabiliste Siméon Denis Poisson « "Lorsqu'il s'agirait de trente boules extraites d'une urne qui contient des nombres égaux de boules blanches et de boules noires, les événements *remarquables*, écrit Poisson, seront l'arrivée de 30 boules de la même couleur, celle de 30 boules alternativement blanches et noires, celle de 15 boules d'une couleur, suivies de 15 boules de l'autre couleur, etc. Dans le cas d'une trentaine de caractères d'imprimerie rangés à la suite l'un de l'autre, les événements *remarquables* seront ceux où ces lettres se trouveront disposées soit dans l'ordre alphabétique, soit dans l'ordre inverse, ou bien ceux où elles formeront une phrase de la langue française, ou d'une autre langue... " en partant de là, remarque Cournot, et en admettant qu'on connaît le nombre des événements *remarquables* et celui de événements *non remarquables*, M. Poisson assigne, d'après les règles communément admises dans la théorie des probabilités *a posteriori*, que l'apparition d'un événement remarquable n'est point au hasard ». La critique de Cournot porte sur le point précis que le fait que la définition exacte de ce qui est remarquable ou non soit délicate, n'est pas une difficulté mineure mais de première importance : « La combinaison de caractères d'imprimerie, dans laquelle un voyageur reconnaîtrait quelques mots parlés par une peuplade sauvage, doit-elle être comptée parmi les combinaisons remarquables, et serait-elle pour nous aussi remarquable que celle qui nous offrirait des mots de notre langue ? »². Cournot montre

¹ Pour plus de détails et des exemples d'hypogrammes cf. N.B. *La règle, le compas et le divan*, Seuil 2001.

² Antoine Augustin Cournot, *Recherches sur les principes mathématiques de la théorie des richesses*, 1838.

que toute la socialité et l'historicité du langage s'engouffrent dans cette question d'une chaîne de caractères *remarquable*.

Il se trouve que Claude Shannon, dans la présentation de sa théorie de l'information, procède à la construction d'un modèle mathématique de la langue anglaise, exactement dans l'esprit de ce que nous nous proposons de faire pour évaluer les découvertes de Saussure, afin de convaincre le lecteur qu'un "processus stochastique suffisamment complexe donne une bonne représentation d'une source discrète". Mais il évince, hélas, le grave problème de préciser les conditions et la méthode pour la "convergence vers la langue anglaise". Doit-on mettre dans le corpus Shakespeare, les textes officiels, la littérature grise, les hypertextes du web ? On a besoin de comprendre pour savoir ce que doit être le pourtour de "la langue anglaise". Lorsque Shannon écrit « Fréquemment les messages ont des *significations* [...] ces aspects sémantiques de la communication ne concernent pas le problème de l'ingénieur. Ce qui importe, c'est que le message actuel est tiré d'un ensemble de messages possibles » il écarte un problème réel qui devient patent lorsqu'on envisage des probabilités sur les mots et sur les phrases.

On ne connaît pas de loi de probabilité privilégiée sur les mots ni sur les phrases de la langue française, ni sur les nombres entiers d'ailleurs, ni sur les molécules chimiquement stables ou sur les protéines.

Le problème est analogue lorsque Jacques Monod parle de "roulette de la nature", il va un peu trop vite car il n'est pas clair que si contingence il y a, elle soit aléatoire ni même qu'elle soit objective et ne traduise pas simplement notre connaissance partielle des mécanismes. Après les critiques fort pertinentes de E. Schoffeniels et G. Matheron, un rôle déterminant doit être accordé au *contexte* que les biologistes appellent *l'épigenèse*, qui *a contrario*, si nous savions le décrire complètement, tendrait à montrer que le monde ne saurait être essentiellement différent de ce qu'il est.

De même, un spécialiste de lexicologie quantitative, observant les théorèmes mathématiques écrits dans le langage des prédicats à partir des axiomes de la théorie des ensembles pourrait considérer qu'ils sont le fruit du hasard parmi les trajets qu'autorise la logique, alors qu'un historien et sociologue des sciences s'appuyant sur le sens des énoncés s'efforcerait de montrer qu'ils résultent de ce qui est déjà su, et des conditions de travail et de préoccupation des chercheurs.

Notons que Jacques Bernoulli (1654-1705) écrivant son ouvrage *Ars conjectandi*, met des probabilités sur les jeux de cartes et de jetons mais se garde bien d'en mettre sur les conjectures mathématiques. Cela ne servirait à rien, car c'est leur sens qui intéresse et c'est avec le sens qu'on manie les mathématiques. Il y a une sorte de *principe d'exclusion* (non de Pauli mais disons de Cournot-Matheron) : *le sens est une catégorie qui exclut le hasard*. On peut mettre des probabilités sur les signifiants (les termes) pas sur les signifiés. Lorsqu'on a trouvé du sens quelque part, on pense que ce n'est pas fortuit. Les découvertes poétiques de Saussure, maintenant qu'elles sont faites, enrichissent définitivement la lecture de ces poèmes et font dorénavant partie constitutive de ces œuvres d'art. De même, lorsque le statisticien, par des corrélations et des analyses en composantes principales sur des données brutes, dégage des structures et des propriétés, il fait apparaître du sens, et est convaincu que cela n'est pas arbitraire mais bien réel. Je pense qu'une des raisons tient à la contextualité et à l'historicité du sens. Une des idées que Jacques Derrida a développées avec la notion de dissémination soulignant que le sens d'un mot échappe même à l'auteur par le fait que le langage est un matériau social qui le dépasse. Nous reviendrons *in fine* sur cette question

L'opinion publique

L'expérience qui m'a fait tiquer sur ce thème est courante et chacun a dû la rencontrer à l'occasion. Il s'agissait d'une étude sur la pédagogie à l'Ecole des Ponts. On avait interrogé par

courrier deux promotions d'anciens élèves sur ce qui leur servait dans la vie professionnelle et ce qu'ils pensaient des divers enseignements. Compte tenu du faible taux de réponses, on avait quarante questionnaires à dépouiller. Il y avait des questions sur les cours de première année, de seconde année, sur les cours optionnels, sur les travaux de fin d'étude, sur les DEA, sur les stages,... Finalement plus de cent questions dont certaines, étant données les choix de parcours par les élèves, n'étaient pertinentes que pour une fraction des quarante réponses. On était en droit de s'interroger sur le bien fondé des conclusions d'une telle étude.

Le mathématicien John Paulos a présenté de façon frappante à quoi pouvait mener l'art d'augmenter les catégories³. Il considère un personnage, en quête de renommée, qui envoie, avant un match de foot, 1024 lettres à des correspondants en prédisant pour 512 d'entre elles que c'est l'équipe A qui va gagner, l'équipe B pour les autres. Puis il recommence à propos des prochaines élections en envoyant 512 lettres à ceux des précédents destinataires pour lesquels il ne s'était pas trompé, en partageant encore en deux ses anticipations... Après quelques itérations, il parvient à constituer un groupe de quelques personnes absolument convaincues qu'il a un remarquable don divinatoire, et qui sont peut-être, sans autres informations, enclines à lui faire confiance une fois encore... Un apprenti gourou est né.

Paulos consacre un chapitre à la question "too many characteristics, not enough people" en montrant comment les corrélations entre caractéristiques peuvent devenir très fausses sur de faibles échantillons.

Le propre des sondages dits "d'opinion" est de poser plusieurs questions par sondages. On lit à ce sujet dans certains manuels de statistiques pour les sciences humaines que "la précision ne dépend pas du nombre de questions posées". Il est certes exact, si le sondage est bien fait, que la loi de probabilité de la réponse à la question n°7 disons, ne dépend pas du nombre de questions posées mais seulement de la taille de l'échantillon et de la proportion véritable dans la population totale. Néanmoins, il est vrai également que si le nombre de questions augmente, la proportion du nombre de caractéristiques pour lesquelles l'échantillon est similaire à la population totale ne s'améliore pas avec le nombre de questions et peut se détériorer sérieusement.

On voit bien le piège quand l'échantillon est de petite taille, quelques dizaines de personnes (comme dans les études dites "qualitatives" nous y reviendrons), si l'on pose plusieurs milliers de questions, on finit sans doute par cerner assez bien les caractéristiques des personnes interrogées, mais on ne tire plus d'information sur la population totale en les questionnant davantage.

Le cas où les caractéristiques sont "indépendantes" et de probabilité 1/2 dans la population totale est sans doute trop souvent cité, car les calculs sont particulièrement aisés (cf. encadré 1). On trouve que la proportion dans l'échantillon du nombre de caractéristiques pour lesquelles les réponses sont correctes à epsilon près converge, lorsque le nombre de question augmente, vers la probabilité que la question n°1 soit correcte à epsilon près dans l'échantillon.

Mais le cas général des sondages avec beaucoup de questions est plus difficile, car les caractéristiques ne sont pas indépendantes (et c'est ce qui nous intéresse souvent : savoir si les fumeurs et les malades du cancer sont corrélés). La situation est aussi générale que celle de la convergence de

$$\frac{1}{K} \sum_{k=1}^K X(k)$$

où $X(k)$ est un processus à temps discret, disons borné, quelconque.

³ John Alan Paulos, *Once upon a number*, Penguin Press, 1998.

Soit une population infinie où la proportion d'individus de caractéristique C est p .

Pour un sondage de n individus
l'écart entre la proportion observée p' et p
suit approximativement une loi normale $\mathcal{N}(0, \frac{p(1-p)}{n})$.

Au seuil de confiance de 95%

$$|p - p'| \leq 1.96 \sqrt{\frac{p(1-p)}{n}}.$$

Avec	$p = \frac{1}{2}$	$ p - p' \leq 0.02$	pour	$n \geq 1950$
------	-------------------	----------------------	------	---------------

Notons $N(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dx$

On a

$ p - p' < a$ au seuil de confiance $b \in]0, 1[$ si $b = 2N(\frac{a}{\sigma}) - 1$ c'est à dire $a = \sigma N^{-1}(\frac{b+1}{2})$

$$\sigma = \sqrt{\frac{p(p-1)}{n}}$$

Si on fait K sondages indépendants concernant chacun une caractéristique différente C_1, \dots, C_K de proportion p_1, \dots, p_K ,

on a

$(p_i - p'_i < a \quad \forall i)$ au seuil de confiance b si $b = (2N(\frac{a}{\sigma}) - 1)^K$ c'est à dire si $a = \sigma N^{-1}(\frac{b^{\frac{1}{K}}+1}{2})$

$$\sigma = \sqrt{\frac{p(p-1)}{n}}$$

Quand on multiplie les sondages, (on en fait depuis le 19^e siècle), la relation entre précision, seuil de vraisemblance, et taille de l'échantillon devient :

Pour qu'au seuil de vraisemblance b , les K sondages aient la précision a , il faut

$\sqrt{n} = \frac{\sqrt{p(1-p)}}{a} N^{-1}(\frac{b^{\frac{1}{K}}+1}{2})$	donc	$n \uparrow \infty$ si $K \uparrow \infty$
--	------	--

Si on multiplie les sondages à $a = 2\%$, $b = 95\%$ et $n = 2000$
il finit par y en avoir de faux, c'est bien évident.

Cas de caractéristiques indépendantes de probabilité $\frac{1}{2}$

On peut considérer que la population est constituée de nombres dyadiques de profondeur K avec la probabilité uniforme, on tire N points de digits binaires $a_{k,n}$.

Si nous disons que l'échantillon est similaire à la population totale pour la caractéristique k lorsque

$$\left| \frac{1}{N} \sum_{n=1}^N a_{k,n} - \frac{1}{2} \right| \leq \frac{1}{4}$$

la question est d'étudier

$$\Pi_K = \frac{1}{K} \sum_{k=1}^K 1_{\left\{ \left| \frac{1}{N} \sum_{n=1}^N a_{k,n} - \frac{1}{2} \right| \leq \frac{1}{4} \right\}}$$

proportion du nombre de question "bien répondues" par l'échantillon

Comme les $a_{k,n}$ sont ici indépendantes pour les divers k , la loi des grands nombres donne :

$$\lim_{K \uparrow \infty} \Pi_K = \mathbb{P} \left\{ \left| \frac{1}{N} \sum_{n=1}^N a_{k,n} - \frac{1}{2} \right| \leq \frac{1}{4} \right\}$$

La proportion du nombre de caractéristiques correctes à ε -près converge vers la probabilité que la caractéristique n°1 soit correcte à ε -près.

Exemples où la situation se dégrade lorsque le nombre de questions augmente

1. caractéristiques discrètes

La population totale est un ensemble fini très grand de nombres rationnels de l'intervalle $]0,1[$.

L'échantillon est pris dedans.

La caractéristique n° k de l'individu x est la k -ième décimale de x .

Comme les décimales des points de l'échantillon sont périodiques l'échantillon lui-même est périodique de période le ppcm des périodes.

Ses décimales lointaines ne renseignent en rien sur le reste de la population.

Il se peut que pour k grand, la répartition des k -ièmes décimales dans la population soit le complément à 1 de celle de l'échantillon.

Exemples où la situation se dégrade lorsque le nombre de questions augmente

2. caractéristiques discrètes se rapprochant du continu

On considère une urne de Polya de configuration initiale : une boule blanche et une boule noire.

On tire une boule dans l'urne et on la remet avec une autre de la même couleur.

On note les proportions de boules noires successivement obtenues

$$p_0 = \frac{1}{2}, p_1(\omega), p_2(\omega), \dots, p_n(\omega), \dots$$

on appelle une telle suite "un jeu".

On peut montrer que $\lim_{k \uparrow \infty} p_k(\omega) = p_\infty(\omega)$ existe p.s. et que la loi de p_∞ est uniforme sur l'intervalle $[0,1]$.

La population totale est un grand nombre de jeux

L'échantillon est un petit nombre de jeux.

La k -ième question est $p_k > \frac{1}{2}$?

On voit que chaque jeu est très différent, et que les propriétés asymptotiques de l'échantillon sont très différentes de celles de la population globale.

Exemples où la situation se dégrade lorsque le nombre de questions augmente

3. caractéristiques continues

Prenons un exemple où chaque individu est une suite infinie de réels:

$$X = (X_1, X_2, \dots, X_k, \dots)$$

définis par

$$X_k = \text{partie fractionnaire de } U^k$$

où U est une variable aléatoire positive fixée mais inconnue.

La k -ième question est $X_k > \frac{1}{2}$?

On se convainc que la connaissance de l'échantillon renseigne très mal sur les propriétés de la population globale.

Les processus stationnaires non-ergodiques fournissent aussi de bons exemples, par exemple

$$X_k = U_1 e^{i\lambda k} + U_2 e^{i\mu k} \quad \lambda, \mu \text{ fixés}$$

Ces phénomènes ont une grande analogie avec la distinction entre convergence faible et convergence forte dans un espace vectoriel de dimension infinie. La difficulté est bien connue en statistique empirique : il est difficile de remplir un espace de grande dimension même avec énormément de données. Certains auteurs ont appelé ce phénomène "the curse of dimensionality", la malédiction de la dimension⁴.

Or les caractéristiques de la population totale sont inconnues et l'attitude la plus honnête est de laisser ouverte les éventualités et d'éviter les conclusions qui résulteraient d'hypothèses *a priori*. Donc pour connaître les opinions, on est amené à poser beaucoup de questions. On sait depuis Condorcet que pour les choix de préférence d'une population entre trois décisions, si l'on ne pose que trois questions (au lieu de $3!=6$), les réponses à la majorité peuvent être contradictoires.

Ici intervient la question philosophique de *l'enquête qualitative* : le questionnement approfondi d'une population restreinte peut elle apporter une connaissance un tant soit peu scientifique ?

Il faut préciser d'emblée que la valeur qu'attachent les sociologues à de telles enquêtes ne tient pas à la croyance à une bonne précision de chaque réponse prise isolément, ni à une hypothèse d'indépendance des propriétés étudiées, mais à d'autres arguments plus intéressants à discuter.

En premier lieu, le sociologue considère que les individus sont sociaux par nombre de leurs caractéristiques (langage, affectivité, communication, etc.) et que le social influence fortement leur conscience donc leur opinion. La phrase célèbre de Marx "ce n'est pas la conscience de l'homme qui détermine son être, mais son être social qui détermine sa conscience" contient une grande part de vérité comme principe méthodologique.

Egalement, l'échantillon peut être utilisé comme "témoin des son temps", comme miroir (cf. l'homme spéculaire de Richard Rorty). On cherche à travers l'enquête qualitative, non la particularité de l'individu interrogé, mais la trace de l'influence sociale dans ses réponses, les contraintes qu'il ressent, les interprétations que ses informations lui suggèrent. C'est la dimension, non pas conformiste de l'individu, mais conformable devrions nous dire, qui est intéressante. D'où fréquemment des questions qui portent sur ses réactions à des faits nouveaux pour lui ou supposés, afin de déceler les points d'appuis sociaux de ses explications.

Les théoriciens de la méthode sociologique vont cependant plus loin : il s'agit par ces enquêtes de faire émerger du non-conscient⁵. «Durkheim, qui exige du sociologue qu'il pénètre dans le monde social comme dans un monde inconnu, reconnaît à Marx le mérite d'avoir rompu avec l'illusion de la transparence : "nous croyons féconde cette idée que la vie sociale doit s'expliquer, non par la conception que s'en font ceux qui y participent, mais par des causes profondes qui échappent à la conscience"»⁶. Bourdieu et ses collaborateurs énoncent un principe de la non-conscience assez fort : "les relations sociales ne sauraient se réduire à des rapports entre des subjectivités animées par des intentions ou des motivations parce qu'elles s'établissent entre des conditions et des positions sociales et qu'elles ont, du même coup, plus de réalité que le sujet qu'elles lient". L'enquête qualitative est un outil pour faire émerger des concepts (des idéaux-types selon Weber) qui n'apparaissaient pas non plus à l'enquêteur *a priori*.

Par exemple une enquête a été réalisée sur les choix d'émissions par les téléspectateurs, les individus de l'échantillon ayant accepté que des dispositifs enregistreurs soient installés sur leur téléviseur pendant un mois. Au dépouillement, on s'est trouvé incapable de mettre en

⁴ Cf. R. E. Bellman, *Adaptive control processes*, Princeton univ. press 1961, et U. Gather et C. Becker, "The curse of dimensionality, a challenge for mathematical statistics", *Jahresbericht der Deut. Math.-Vereinigung*, 103, 19-36, 2001.

⁵ Cf. A. Sauvy, *L'opinion publique*, Que sais-je PUF 1971

⁶ P. Bourdieu, J.-Cl. Chamboredon, J.-Cl. Passeron, *Le métier de sociologue*, Mouton 1983.

correspondance les choix et des types d'émissions. Une enquête qualitative ultérieure révéla que nombre de téléspectateurs "zapaient", c'est-à-dire passaient en revue toutes les chaînes jusqu'à obtenir des choses plaisantes selon des critères beaucoup plus variés que les "types d'émissions", ceux-ci apparaissant dès lors par trop réducteurs.

Les précautions méthodologiques des sociologues, balancées avec leur souci évident de maintenir la fécondité de leur discipline, ne nous renseignent pas finalement sur la pertinence du concept d'opinion publique. Alfred Sauvy la définit comme

- une force politique
- qui doit rencontrer une certaine résistance pour pouvoir s'exercer, également un point d'appui
- qui n'est pas nécessairement la résultante d'opinions individuelles ni même de la majorité sur un thème donné
- qui présente des positions permanentes et des variations, les courants d'opinion.

Alfred Sauvy parle d'opinion publique nationale et même mondiale.

Forçons le trait. L'homme sociologique est une manifestation observable du social, grâce à laquelle le sociologue propose des interprétations. le sujet-individu ainsi étudié a "moins de réalité" que les relations sociales. Par ailleurs, l'homme économique est une sorte d'automate randomisé suivant des règles de rationalité ou de rationalité limitée (hypothèse d'individualisme méthodologique) le but des modélisations étant de rendre compte avec de tels agents des phénomènes sociaux observés.

Dans les deux cas, le vêtement est un peu étroit. Il semble qu'on a, sinon omis, du moins sous-estimé la dimension langagière de l'humain. Les symbolismes de langage agissent sur le sujet et le sujet agit et *se constitue* par le langage. Ce constat apporte une limitation épistémologique fondamentale à ces deux disciplines. En croyant comprendre, l'économiste et le sociologue font eux-mêmes opérer une dimension langagière dont ils surestiment la portée qui reste strictement disciplinaire.

L'infinitude de l'individu

Essayons de donner la tournure la plus concrète à notre discussion. Dans quels cas parle-t-on de dimension infinie en physique ou en mathématique ? Lorsqu'on rencontre des paramètres "répartis", autrement dit des fonctions arbitraires ou pouvant varier de façon plus générale que par une description analytique à l'aide d'un nombre fini de nombres réels. On rencontre de telles situations en mécanique des milieux continus, pour les écoulements, en fait dans un grand nombre de modélisations rencontrées par l'ingénieur aujourd'hui grâce au langage de l'analyse fonctionnelle. Les fonctions y sont pensées comme des points d'un espace abstrait, ont des coordonnées, et possèdent des développements en série sur des fonctions de base. Les situations infini-dimensionnelles ne peuvent se traiter que par *approximation*, et c'est la puissance de l'analyse fonctionnelle de fournir des méthodes de résolution approchées (éléments finis, etc.).

A l'opposé philosophiquement de ce type de représentation, sont d'une part les situations paramétrées analogue à la mécanique d'un mobile dans l'espace et d'autre part les situations discrètes régies par des nombres entiers et des combinaisons rigoureuses, sans approximation. Les notions d'automate et d'algorithme (machine de Turing) sont pris en logique en ce sens strictement combinatoire.

Placer l'individu humain dans la seconde catégorie signifie que l'on considère les savoirs scientifiques, qui sont sociaux, cumulatifs, d'ores et déjà considérables et en développement, susceptibles de faire le tour de tels systèmes. C'est d'ailleurs ce type d'argument qu'emploie J.-

P. Changeux dans son dialogue avec Alain Connes pour esquiver la limitation de réflexivité et ouvrir le champ le plus large à l'investigation du cerveau humain⁷.

Les découvertes du vingtième siècle sur le code génétique viennent conforter cette position. Les caractéristiques des ADN sont complexes mais finies. Mettre en correspondance les phénotypes avec les génotypes serait au fond un problème d'évaluation de la sensibilité aux conditions initiales. La puissance de l'informatique, la vitesse de son développement et les perspectives mirifiques d'ordinateurs utilisant les propriétés de certaines molécules organiques ou des propriétés quantiques, contrebalance les arguments des spécialistes de la complexité algorithmique et les limitations effectives des problèmes NP-complets et autres.

L'autre choix est une position philosophique en soi. Elle revient à mettre l'accent sur ce qui n'est pas réductible à des représentations systémiques, à attacher de la valeur à ce qui fait sortir l'individu des catégories de pensée pré-établies. Nous retrouvons ici l'importance de l'épigénèse, l'ADN fournit des protéines dans un milieu cytoplasmique, la cellule se scinde dans un tissu, l'embryon se développe dans la poche placentaire ; l'enfance, la famille, les mots, l'école, fabriquent un sujet dont l'histoire est constitutive de son être. Cette position n'est certes pas une théorie philosophique charpentée, simplement une thèse qui veut accorder tout le poids aux potentialités individuelles favorisées par l'éducation, la culture.

La science simplifie et chaque discipline plaide toujours pour ses propres exploits futurs. Mais elle est faite par nous, et c'est à nous de l'orienter de telle sorte que les valeurs de civilisation soient préservées.

⁷ *Matière à pensée*, Odile Jacob 1999.